A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings

Biao Wang, *Student Member, IEEE*, Yaguo Lei, *Member, IEEE*, Naipeng Li, and Ningbo Li

***Abstract*—Remaining useful life (RUL) prediction of rolling element bearings plays a pivotal role in reducing costly unplanned maintenance and increasing the reliability, availability, and safety of machines. This paper proposes a hybrid prognostics approach for RUL prediction of rolling element bearings. First, degradation data of bearings are sparsely represented using relevance vector machine regressions with different kernel parameters. Then, exponential degradation models coupled with the Fréchet distance are employed to estimate the RUL adaptively. The proposed approach is evaluated using the vibration data from accelerated degradation tests of rolling element bearings and the public PRONOSTIA bearing datasets. Experimental results demonstrate the effectiveness of the proposed approach in improving the accuracy and convergence of RUL prediction of rolling element bearings.**

***Index Terms*—Bearing degradation, prognostics, relevance vector machine, remaining useful life estimation, vibration monitoring.**

NOMENCLATURE

Acronyms

RUL	Remaining useful life.
ABMA	American Bearing Manufacturers Association.
CM	Condition monitoring.
ML	Machine learning.
ANNs	Artificial neural networks.
SVM	Support vector machine.
RVM	Relevance vector machine.
DBN	Deep belief network.
KF	Kalman filtering.
EKF	Extended Kalman filtering.
PF	Particle filtering.
RVs	Relevance vectors.
MA	Maximum amplitude.
CRA	Cumulative relative accuracy.
SD	Standard deviation.

Notation

$\{\boldsymbol{x}_n, z_n\}_{n=1}^N$	Input-target pairs of RVM regression model.
$y(\boldsymbol{x}_n)$	Linear combination of kernel functions.
ε_n	Additive noise component corresponding to the nth input-target pair.
σ^2	Variance of Gaussian distribution.
$\boldsymbol{w}$	Weight vector of the linear model $y(\boldsymbol{x}_n)$.
$K(\cdot,\cdot)$	Kernel function.
$\boldsymbol{\Phi}$	$N \times N$ Kernel matrix.
$\boldsymbol{\alpha}$	Hyperparameter vector in RVM regression model.
$\{f_n\}_{n=1}^N, \{t_n\}_{n=1}^N$	Extracted features and their corresponding sampling time.
$\{T_k\}_{k=1}^K$	Inspection time.
f_m^*	Predictive mean of the feature extracted at sampling time t_m.
$f_{m_{\text{upper}}}^*, f_{m_{\text{lower}}}^*$	Upper and lower confidence limits of f_m^*.
S	Smoothed degradation trajectory.
$\{F_l\}_{l=1}^L$	Fitted curves by exponential degradation models.
$\delta(S, F_l)$	Fréchet distance between S and F_l.
$\xi(\cdot), \psi(\cdot)$	Two arbitrary continuous nondecreasing functions with $\xi(0) = \psi(0) = t_1$ and $\xi(1) = \psi(1) = t_N$.
$D(\cdot,\cdot)$	Euclidean distance.
ω	Predefined failure threshold.
$\mathrm{RUL}(T_k)$	Estimated RUL at inspection time T_k.
$\mathrm{ActRUL}(T_k)$	Actual RUL at inspection time T_k.
$\mathrm{RA}(T_k)$	Relative prediction accuracy at inspection time T_k.
C_{PE}	Convergence of RUL prediction results.
(C_x, C_y)	Centroid of the area under the prediction error curve.
Er_i	Percent error of RUL prediction result for ith testing dataset.

Manuscript received June 19, 2018; revised September 19, 2018; accepted November 17, 2018. Date of publication December 14, 2018; date of current version March 2, 2020. This work was supported in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1709208, in part by the National Natural Science Foundation of China under Grant 61673311 and Grant 51421004, and in part by the National Program for Support of Top-Notch Young Professionals. Associate Editor: R. Kang. *(Corresponding author: Yaguo Lei.)*

The authors are with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wangbiaoxjtu@outlook.com; yaguolei@mail.xjtu.edu.cn; li3112001096@stu.xjtu.edu.cn; liningbo1992@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TR.2018.2882682

I. INTRODUCTION

NOWADAYS, rolling element bearings are widely used in various rotary machines, such as drills, electric motors, wind turbines, and turbofan engines. During machine operation, different types of faults often occur in bearings due to overload, poor lubrication, improper mounting, etc., [1]–[4]. If no effective action is taken in time, bearing failures may cause the breakdown of the entire machine and even catastrophes. In order to avoid that, remaining useful life (RUL) prediction of rolling element bearings has emerged as a critical technology to provide failure warnings in advance and improve maintenance schedules, which reduces costly unplanned maintenance and enhances the reliability, availability, and safety of machines [5].

The RUL of bearings is defined as the time span from the current inspection time to the future time instant at which the bearing will no longer perform its intended function. Traditionally, the lifetime of a bearing can be calculated based on the international standard ISO 281-2007 or the American Bearing Manufacturers Association (ABMA) standards ABMA 9-2015 and ABMA 11-2014 [6]. The actual lifetime of the bearing, however, is significantly different from the theoretical one owing to severe operating conditions, manufacturing defects, etc. Current approaches for RUL prediction of bearings can be divided into two major categories: data-driven approaches and model-based approaches [7]–[9].

Data-driven approaches use event data and condition monitoring (CM) data coupled with machine learning (ML) techniques, e.g., artificial neural networks (ANNs) [10], [11], support vector machine (SVM) [12], [13], and relevance vector machine (RVM) [14], [15], to train a prediction model and then utilize the trained model to estimate the RUL of a bearing. The advantage of data-driven approaches is that they can directly learn the underlying degradation trends of bearings from the available sensor data, and hence, the users do not need to know the exact failure mechanisms of bearings. Soualhi *et al.* [16] used the adaptive neurofuzzy inference system to estimate RUL of bearings. Ali *et al.* [17] developed a prognostics methodology based on ANNs for accurate bearing RUL prediction. Javed *et al.* [18] proposed a summation wavelet extreme learning machine approach to perform the RUL estimation task of bearings. However, to acquire accurate prediction results, data-driven approaches require a considerable number of run-to-failure data to train the prediction model [8], which may be cost prohibitive or impractical in real cases. Especially, since no physical knowledge is involved in the data-driven approaches, the degradation prediction process is usually opaque to users and the results may be counter-intuitive [7]. In addition, the performance of a data-driven approach also largely depends on the parameter selection of the adopted ML model [19]. For example, the selection of kernel functions substantially influences the accuracy and generalization performance of the RVM-based prognostics approaches [14].

Model-based approaches, on the other hand, use mathematical or physical models derived from first principles and the knowledge of failure mechanisms to describe the degradation processes of bearings. Then, statistical estimation techniques, such as Kalman filtering (KF) [20], [21], extended KF (EKF) [6], [22], and particle filtering (PF) [23], [24], are employed to identify and update model parameters and predict RUL of bearings. Liao [23] used the Pairs–Erdogan model to describe the fault growth of bearings, and predicted the RUL using a continuous Bayesian updating approach. Bolander *et al.* [25] developed a spall propagation model for aircraft engine bearings, and used a PF-based approach to predict the spall propagation rate. Li *et al.* [24] employed an exponential degradation model to describe the degradation processes of bearings, and estimated the RUL with the help of the PF algorithm. The main merits of model-based approaches are that the prediction results tend to be intuitive and the computational efficiency is higher than data-driven approaches. Model-based approaches may work well when the degradation processes of bearings can be accurately described using the developed degradation models [26]–[28]. In industrial applications, the failure mechanisms of bearings, however, are generally various or not straightforward, which means that accurate degradation models may be unavailable [29]. Moreover, a model-based approach is often developed based on a specific case and also requires explicit prior knowledge or extensive empirical data to initialize model parameters [4]. Therefore, model-based approaches have limited applications in RUL prediction of bearings.

To cope with the aforementioned limitations in both data-driven and model-based approaches, a hybrid prognostics approach that combines RVM regressions, exponential degradation models, and Fréchet distance is proposed to predict RUL of rolling element bearings in this paper. In the proposed hybrid prognostics approach, RVM regressions with different kernel parameters are first used to sparsely represent the features extracted from real-time degradation signals of bearings. Then, exponential degradation models coupled with the Fréchet distance are employed to estimate the RUL adaptively. Through incorporating CM data (i.e., extracted feature), ML technique (i.e., RVM regression), and prior model (i.e., exponential degradation model) into RUL prediction of bearings, the proposed approach not only integrates the strengths from both data-driven and model-based approaches but also alleviates their limitations. On one hand, the employment of exponential degradation models reduces the dependence on historical data and can help to predict the future degradation trends of bearings more accurately, especially at the initial prediction stage where only a few degradation data are available. On the other hand, in place of initializing model parameters based on empirical knowledge and updating them using statistical techniques, the proposed approach directly uses different relevance vectors (RVs) obtained by RVM regressions to identify the unknown model parameters, and then utilizes the Fréchet distance to adaptively find the optimal one, thus effectively improving the accuracy and convergence of RUL estimation of bearings.

The main contributions of this paper are summarized as follows.

1) RVM regressions with different kernel parameters are used to obtain different representative data from the available degradation data of bearings. This processing not only avoids the problem of kernel parameter selection but also inherits the advantages of RVM, e.g., filtering the unwanted measurement noise and managing the uncertainty in prognostics. More importantly, this processing takes full advantage of the sparsity of RVM so as to find various representative degradation information from the limited measurements.
2) The Fréchet distance is introduced into the field of RUL prediction, which is a new attempt in this field. With the

help of the Fréchet distance, the problem of model parameter initialization is alleviated effectively. Meanwhile, the combination of the Fréchet distance and the exponential degradation models is able to better describe the possible degradation paths of a bearing and so the degradation predictions are more accurate.

3) A hybrid prognostics strategy is developed by systematically fusing RVM regressions, exponential degradation models, and Fréchet distance, which is able to provide more accurate RUL estimations based on real-time degradation data of bearings, rather than relying on the building of a precise degradation model or the use of a large number of historical data.

The rest of this paper is organized as follows. Section II briefly introduces the basic theory of RVM regression. Section III details the proposed hybrid prognostics approach. In Sections IV and V, the vibration data from the accelerated degradation tests of rolling element bearings and the public PRONOSTIA bearing datasets are used to demonstrate the effectiveness and superiority of the proposed approach, respectively. Finally, conclusions are drawn in Section VI.

II. Basic Theory of RVM Regression

RVM, one of the commonly used ML techniques, is a probabilistic sparse kernel model based on the Bayesian framework. It offers excellent sparsity and good generalization performance in regression and classification. In this section, the basic theory of RVM for regression analysis is first introduced, and then its sparsity is discussed.

A. RVM Regression

Given a dataset of input-target pairs $\{\boldsymbol{x}_n, z_n\}_{n=1}^{N}$, RVM regression describes the relationship between targets and inputs using a linear model of the form

$$z_n = y\left(\boldsymbol{x}_n\right) + \varepsilon_n \tag{1}$$

where $y(\boldsymbol{x}_n)$ is a linear combination of kernel functions, and $\varepsilon_n = \sigma^2$ is an additive noise component corresponding to the nth input-target pair. The mathematical expression of $y(\boldsymbol{x}_n)$ is given by

$$y\left(\mathbf{x}_n\right) = \sum_{i=1}^{N} w_i K\left(\boldsymbol{x}_i, \boldsymbol{x}_n\right) = \boldsymbol{w}^{\mathrm{T}} K\left(\boldsymbol{x}, \boldsymbol{x}_n\right) \tag{2}$$

where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^{\mathrm{T}}$, $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^{\mathrm{T}}$ is the weight vector, and $K(\cdot,\cdot)$ is a predetermined kernel function. Thus, the conditional probability for target variable z_n is a Gaussian distribution with mean $y(\boldsymbol{x}_n)$ and variance σ^2, and then the likelihood of the target vectors can be written as

$$\begin{aligned} p\left(\boldsymbol{z}\left|\boldsymbol{w}, \sigma^2\right.\right) &= \prod_{n=1}^{N} p\left(z_n \left|\boldsymbol{w}, \sigma^2\right.\right) = \left(2\pi\sigma^2\right)^{-N/2} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\left\|\boldsymbol{z} - \boldsymbol{\Phi}\boldsymbol{w}\right\|^2\right\} \end{aligned} \tag{3}$$

where $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)^{\mathrm{T}}$ and $\boldsymbol{\Phi}$ is a $N \times N$ kernel matrix with $\boldsymbol{\Phi} = [K(\boldsymbol{x}, \boldsymbol{x}_1), K(\boldsymbol{x}, \boldsymbol{x}_2), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_N)]^{\mathrm{T}}$.

The estimation of $\boldsymbol{w}$ and σ^2 in (3) using the maximum likelihood estimation method often suffers from overfitting. To avoid that, an automatic relevance determination Gaussian prior over $\boldsymbol{w}$ [30] is introduced

$$p\left(\boldsymbol{w}\left|\boldsymbol{\alpha}\right.\right) = \prod_{i=1}^{N} N\left(w_i\left|0, \alpha_i^{-1}\right.\right) \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)^{\mathrm{T}}$ is a hyperparameter vector and each individual hyperparameter α_i represents the precision of the corresponding weight w_i. Using Bayes' rule, the posterior distribution of $\boldsymbol{w}$ is given by

$$p\left(\boldsymbol{w}\left|\boldsymbol{z}, \boldsymbol{\alpha}, \sigma^2\right.\right) = \frac{p\left(\boldsymbol{z}\left|\boldsymbol{w}, \sigma^2\right.\right) p\left(\boldsymbol{w}\left|\boldsymbol{\alpha}\right.\right)}{p\left(\boldsymbol{z}\left|\boldsymbol{\alpha}, \sigma^2\right.\right)} \tag{5}$$

and follows the Gaussian distribution $N(\boldsymbol{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{z} \tag{6}$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Phi} + \boldsymbol{A}\right)^{-1} \tag{7}$$

where $\boldsymbol{A} = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_N)$ and $\boldsymbol{B} = \sigma^{-2}\boldsymbol{I}_N$. Note that σ^2 is also treated as a hyperparameter, which is optimized together with $\boldsymbol{\alpha}$. By integrating out the model parameters, the logarithm marginal likelihood of $\boldsymbol{\alpha}$ and σ^2 can be obtained

$$\begin{aligned} \ln p\left(\boldsymbol{z}\left|\boldsymbol{\alpha}, \sigma^2\right.\right) &= \ln\int p\left(\boldsymbol{z}\left|\boldsymbol{w}, \sigma^2\right.\right) p\left(\boldsymbol{w}\left|\boldsymbol{\alpha}\right.\right) \mathrm{d}\boldsymbol{w} \\ &= -\frac{1}{2}\left[N\ln\left(2\pi\right) + \ln\left|\boldsymbol{C}\right| + \boldsymbol{z}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{z}\right] \end{aligned} \tag{8}$$

with

$$\boldsymbol{C} = \sigma^2\boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}. \tag{9}$$

The values of the hyperparameters $\boldsymbol{\alpha}$ and σ^2 that maximize $\ln p(\boldsymbol{z}|\boldsymbol{\alpha}, \sigma^2)$ cannot be got in a closed form. In general, the hyperparameters $\boldsymbol{\alpha}$ and σ^2 are optimized using the type-II maximum likelihood method or expectation maximization algorithm. However, these methods repeatedly compute and invert a Hessian matrix of size $N \times N$, which requires $O(N^2)$ memory storage and $O(N^3)$ computation. Therefore, the optimization procedure is very time consuming and memory consuming. In order to reduce the computational burden, this paper uses a faster marginal likelihood maximization algorithm, namely the sequential sparse Bayesian learning algorithm [31]. Compared to the type-II maximum likelihood algorithm or expectation maximization algorithm, the sequential sparse Bayesian learning algorithm requires less memory storage and computation, time therefore, resulting in a very clear speed advantage and decreasing the computational cost.

After obtaining the optimal hyperparameter values $\boldsymbol{\alpha}_{MP}$ and σ_{MP}^2, the predictive distribution over z_* for a new input vector

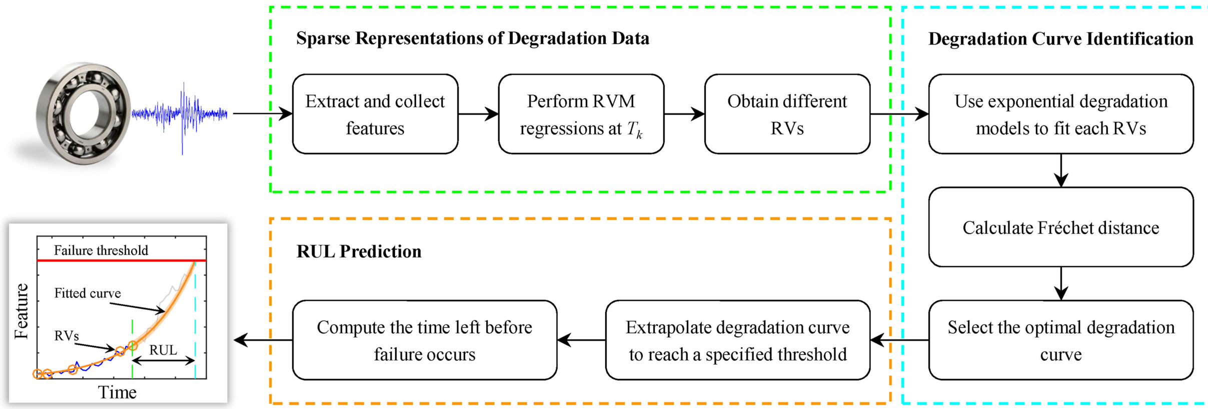


Fig. 1. Flowchart of the proposed hybrid prognostics approach.

$\boldsymbol{x}_*$ can be calculated by

$$\begin{aligned} p\left(z_* \left| \boldsymbol{z}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2\right.\right) &= \int p\left(z_* \left| \boldsymbol{w}, \sigma_{MP}^2\right.\right) p\left(\boldsymbol{w} \left| \boldsymbol{z}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2\right.\right) \mathrm{d}\boldsymbol{w} \\ &= N\Big(z_* \left| \boldsymbol{\mu}_{MP}^{\mathrm{T}} K\left(\boldsymbol{x}, \boldsymbol{x}_*\right), \sigma_{MP}^2 + K(\boldsymbol{x}, \boldsymbol{x}_*)^{\mathrm{T}} \right. \\ &\qquad \times \boldsymbol{\Sigma}_{MP} K\left(\boldsymbol{x}, \boldsymbol{x}_*\right)\Big). \end{aligned} \tag{10}$$

B. Sparsity

During the hyperparameter optimization, many of $\boldsymbol{\alpha}$ values tend to be infinite, so the weights corresponding to these hyperparameters have posterior distributions with mean and variance both zero. This means that those parameters and their corresponding kernel functions play no role in making regression analysis and can be pruned from the model, and then the sparsity of RVM regression is achieved. Furthermore, considering the mechanism of automatic relevance determination introduced in (3), the inputs corresponding to the remaining nonzero weights are called RVs.

According to (3), it can be seen that the sparsity of RVM regression is highly dependent upon the choice of kernel functions (types and parameters). Although scholars have developed some kernel optimization methods based on the metaheuristic algorithms, such as particle swarm optimization, ant colony optimization, and genetic algorithm, no standard or general methods can be found in the literature. Thus, it is still a challenge to find an appropriate kernel function in RVM regression [7]. In this paper, the main purpose of performing RVM regression is to obtain different RVs from the available degradation data of bearings, which means that it is needless to select or design a special kernel function or to find an optimal kernel parameter. Therefore, the Gaussian kernel, which is the most commonly used kernel function, is utilized in this paper, but its parameter is variable for getting different RVs. The Gaussian kernel is defined as

$$K\left(\boldsymbol{x}, \boldsymbol{x}_n\right) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_n\|^2}{2\gamma^2}\right) \tag{11}$$

where γ is the kernel parameter, also named the kernel width. Particularly, in order to avoid overfitting and oversmoothing, the value range of γ is set to [0.5, 50] with an interval of 0.5.

III. Proposed Hybrid Prognostics Approach

The flowchart of the proposed hybrid prognostics approach is shown in Fig. 1. The features that indicate the degradation states of bearings are first extracted from real-time degradation signals. Then, at each inspection time, RVM regressions with different kernel parameters are performed on the extracted features so as to obtain different RVs. Next, two exponential degradation models, i.e., a single exponential function and the sum of two exponential functions, are used to fit these RVs, and the Fréchet distance is employed to select the optimal degradation curve from all fitted curves. Finally, the RUL is estimated by extrapolating the degradation curve to reach a specified failure threshold. The details of this approach are addressed in the following sections.

A. Sparse Representations of Degradation Data

The degradation information of bearings is usually hidden in raw degradation signals, such as vibration or acoustic emission signals. Therefore, in order to reflect the failure progression of bearings, the feature that reflects the degree of bearing degradation is first extracted from raw degradation signals, such as root mean square, kurtosis, or crest factor [10]. Then, the extracted features $\{f_n\}_{n=1}^N$ and their corresponding sampling time $\{t_n\}_{n=1}^N$ are collected as a set for predicting the RUL of the bearing at inspection time $\{T_k\}_{k=1}^K$.

At each inspection time T_k, RVM regressions with different values of kernel parameter are first implemented to sparsely represent the available degradation data $\{t_n, f_n\}_{n=1}^N$, i.e., to obtain different RVs. Each of the obtained RVs consists of the most representative data pairs $\{t_m, f_m^*\}_{m=1}^M$ corresponding to the nonzero weights. Here, f_m^* is the predictive mean of the feature extracted at sampling time t_m, and can be analytically computed by (10), i.e., $f_m^* = \boldsymbol{\mu}_{MP}^{\mathrm{T}} K(\boldsymbol{t}, t_m)$, where $\boldsymbol{t} = (t_1, t_2, \ldots, t_M)^{\mathrm{T}}$. Additionally, since the predictive distribution over the extracted features is Gaussian with variance

$\sigma^2_{MP} + K(\boldsymbol{t}, t_m)^{\mathrm{T}} \boldsymbol{\Sigma}_{MP} K(\boldsymbol{t}, t_m)$, the 95% upper and lower confidence limits of f^*_m can be calculated by the following equations:

$$f^*_{m_{\text{upper}}} = f^*_m + 1.96\sqrt{\sigma^2_{MP} + K(\mathrm{t}, t_m)^{\mathrm{T}} \boldsymbol{\Sigma}_{MP} K(\boldsymbol{t}, t_m)} \tag{12}$$

$$f^*_{m_{\text{lower}}} = f^*_m - 1.96\sqrt{\sigma^2_{MP} + K(\boldsymbol{t}, t_m)^{\mathrm{T}} \boldsymbol{\Sigma}_{MP} K(\boldsymbol{t}, t_m)}. \tag{13}$$

B. Degradation Curve Identification

After different RVs are obtained by RVM regressions, exponential degradation models are used to fit these RVs, aiming to predict the future degradation states of the bearing. Generally, the exponential degradation model is given by a single exponential function. However, in recent years, the sum of two exponential functions [32] has received major attention in the field of RUL prediction. In particular, some scholars have demonstrated the suitability of the sum of two exponential functions in depicting bearing degradation [22]. Thus, in order to better describe the degradation processes of the bearing, this paper uses the single exponential function and the sum of two exponential functions to fit the obtained RVs. The mathematical expressions of the single exponential function and the sum of two exponential functions are given as follows:

$$\begin{cases} f(t) = a\exp(bt) + c \\ f(t) = a\exp(bt) + c\exp(dt) \end{cases} \tag{14}$$

where f is the extracted feature, t is the sampling time, and a, b, c, and d are four unknown parameters that can be estimated using curve fitting methods. In this paper, the nonlinear least square [33] is employed to estimate these unknown parameters.

By fitting each of the obtained RVs, the unknown parameters in (14) can be uniquely identified, and so different degradation curves are obtained at inspection time T_k. However, these fitted curves characterize different bearing degradation trends. Accordingly, the optimal degradation curve, whose trend is closest to the true degradation trend of the bearing, should be selected from all fitted curves and then used to estimate the future degradation states of the bearing. At inspection time T_k, the bearing degradation trend can be reflected by the degradation trajectory, i.e., the feature curve. But the extracted features contain a great amount of measurement noise, which may lead to random outliers or spurious fluctuations. Therefore, in order to filter unwanted measurement noise and capture the true degradation trend of the bearing, it is necessary to perform a smoothing process on the extracted features. Here, an LOESS filter [34] with a span value of 0.3 is used to smooth the extracted features.

Given the smoothed degradation trajectory S and all fitted curves $\{F_l\}_{l=1}^L$, this paper employs the Fréchet distance [35] to find the optimal degradation curve. The Fréchet distance is a superior measure of similarity between curves. In comparison with other distance measure methods, e.g., Euclidean distance, Mahalanobis distance, and Hausdorff distance, the Fréchet distance takes the location and ordering of the points along the curves into account and, thus, is more suitable for the similarity measure between time series curves [36]. Given two curves, i.e., the smoothed degradation trajectory S and one of the fitted curves F_l, the Fréchet distance $\delta(S, F_l)$ is calculated by

$$\delta(S, F_l) = \inf_{\xi,\psi} \max_{x\in[0,1]} \{D(S(\xi(x)), F_l(\psi(x)))\} \tag{15}$$

where $\xi(\cdot)$ and $\psi(\cdot)$ are two arbitrary continuous nondecreasing functions with $\xi(0) = \psi(0) = t_1$ and $\xi(1) = \psi(1) = t_N$, and $D(\cdot,\cdot)$ represents the Euclidean distance. The smaller the Fréchet distance is, the more similar the two curves are. Therefore, the fitted curve whose Fréchet distance is smallest among all of the fitted curves is selected as the optimal degradation curve and then used to predict the degradation states of the bearing at the future time steps.

C. RUL Prediction

At inspection time T_k, the future degradation states of the bearing can be predicted by extrapolating the selected optimal degradation curve. In addition, the uncertainty boundaries of the prediction can also be provided by using (14) to fit the corresponding data pairs $\{t_m, f^*_{m_{\text{upper}}}\}_{m=1}^M$ and $\{t_m, f^*_{m_{\text{lower}}}\}_{m=1}^M$. When the predictive degradation state reaches a specified failure threshold for the first time, the bearing is deemed to be nonoperable. Therefore, according to the concept of the first hitting time [37], the RUL of the bearing can be defined as

$$\mathrm{RUL}(T_k) = \inf\{r : f(r + T_k) \geq \omega \,|\, \boldsymbol{f}\} \tag{16}$$

where $\mathrm{RUL}(T_k)$ is the RUL at T_k, $f(r + T_k)$ is the predictive degradation state at $r + T_k$, $\boldsymbol{f}$ is all of the extracted features at T_k, and ω is the failure threshold, which is predefined by the industrial standards or the operational experiences.

IV. Case Study I: Accelerated Degradation Experiments of Rolling Element Bearings

In this section, the run-to-failure data acquired from accelerated degradation tests of rolling element bearings are used to demonstrate the effectiveness of the proposed hybrid prognostics approach. Furthermore, the proposed approach is also compared with four state-of-the-art prognostics approaches, including two purely data-driven approaches and two purely model-based approaches.

A. Data Description

The bearing testbed is shown in Fig. 2. This platform is able to conduct accelerated degradation tests of bearings to provide real experimental data that characterize the degradation of bearings during the whole operating life. As tabulated in Table I, fifteen rolling element bearings whose type is LDK UER204 are tested under three different operating conditions. Fig. 3 displays the photos of normal and degraded bearings. It can be seen that failure modes of the tested bearings are various, including inner race wear, outer race wear, outer race fracture, etc. To acquire the run-to-failure data of the tested bearings, as shown in Fig. 2, two PCB 352C33 accelerometers are placed on the housing of the tested bearings and positioned at 90° to each other, i.e., one

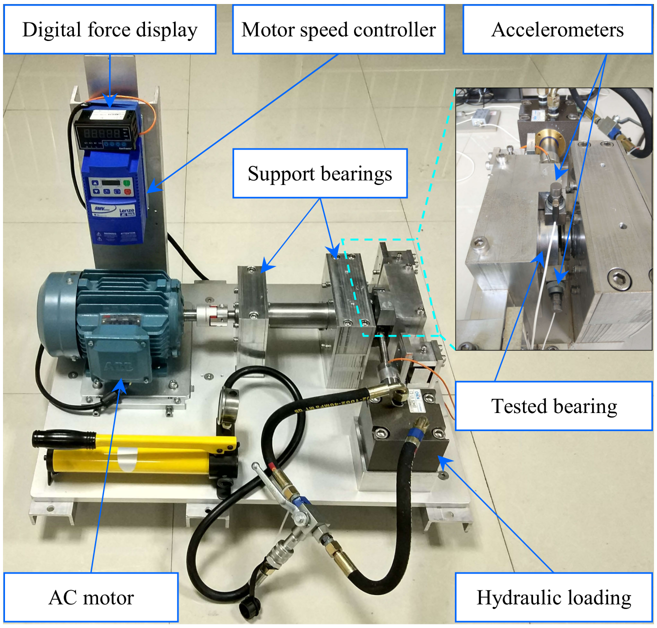


Fig. 2. Bearing testbed.

TABLE I
OPERATING CONDITIONS OF THE TESTED BEARINGS

Operating condition	Radial force (kN)	Rotating speed (rpm)	Bearing dataset
Condition 1	12	2100	Bearing 1_1 Bearing 1_2 Bearing 1_3 Bearing 1_4 Bearing 1_5
Condition 2	11	2250	Bearing 2_1 Bearing 2_2 Bearing 2_3 Bearing 2_4 Bearing 2_5
Condition 3	10	2400	Bearing 3_1 Bearing 3_2 Bearing 3_3 Bearing 3_4 Bearing 3_5

Fig. 3. Photos of tested bearings. (a) Normal bearing. (b) Inner race wear. (c) Outer race wear. (d) Outer race fracture.

is placed on the vertical axis and the other one is placed on the horizontal axis. The sampling frequency is 25.6 kHz, and 32 768 samples (i.e., 1.28 s) are recorded every 1 min. Because the load is applied in the horizontal direction, the accelerometer placed in this direction is able to capture more degradation information of the tested bearings. Therefore, the horizontal vibration signals are selected to estimate the RUL of the tested bearings.

Fig. 4 shows the typical horizontal vibration signals during the whole operating life. It can be seen that the complete bearing degradation process is comprised of two distinctly different stages, i.e., the normal operating stage and the degradation stage. In the normal operating stage, the vibration signals only have random fluctuations at a low level, and so it is difficult to find any degradation information of the tested bearings from the vibration signals. Whereas in the degradation stage, the amplitudes of vibration signals present an increasing trend over operating time, which means that the vibration signals in this stage contain rich information about the bearing degradation. Therefore, the RUL prediction is carried out when the tested bearings start to deteriorate. The beginning point of the degradation stage is detected by employing the adaptive degradation detection method in [24]. In addition, the accelerated degradation tests of bearings are stopped when the amplitude of the vibration signal is higher than 20 g for security reasons. Correspondingly, the time when the amplitude of the vibration signal exceeds 20 g is considered as the failure time of the tested bearing.

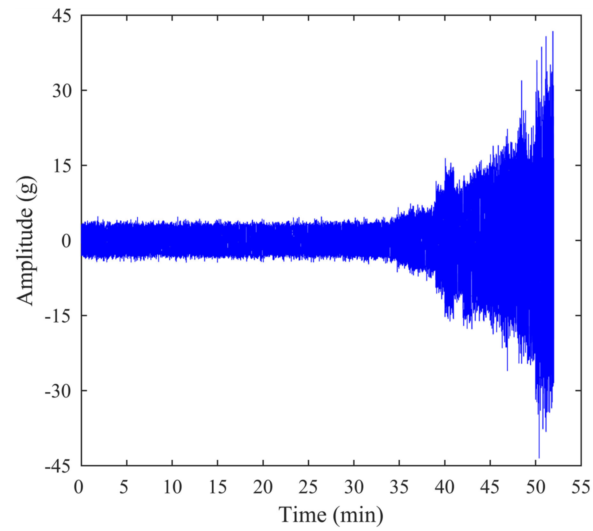


Fig. 4. Typical horizontal vibration signals.

B. RUL Prediction for the Tested Bearings

The maximum amplitude (MA), i.e., the maximum of samples acquired at each sampling time, is first extracted from raw vibration signals to monitor failure progressions of the tested bearings. Then, after the degradation point is detected, the developed hybrid prognostics approach is applied to predicting the future degradation states of the tested bearings and estimating the RUL at each inspection time. To describe the prognostics procedure in detail, Fig. 5 plots the complete degradation prediction process of Bearing 3_1 at inspection time $T_{14} = 2445\,\mathrm{min}$. First of all, the available MA values, which are shown by the blue solid lines in Fig. 5, are input into RVM to perform regression analyses by setting different Gaussian kernel parameter values. Next, both the single exponential function and the sum of two exponential functions in (14) are used to fit each RVs obtained by RVM regressions. Through calculating the Fréchet distance between the smoothed trajectory and each of the fitted

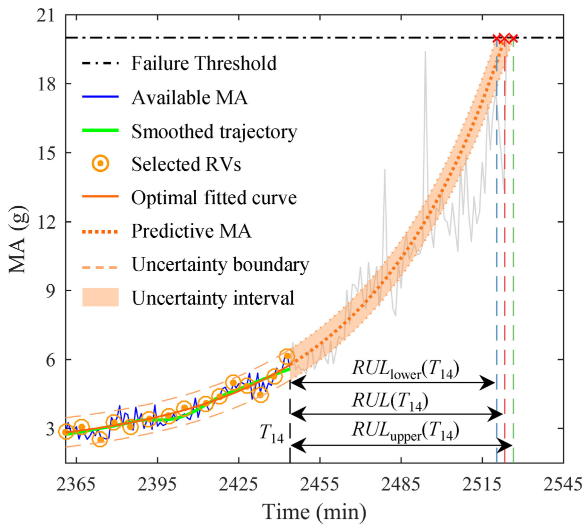


Fig. 5. Illustration of degradation prediction process using Bearing 3_1 at inspection time $T_{14} = 2445$ min.

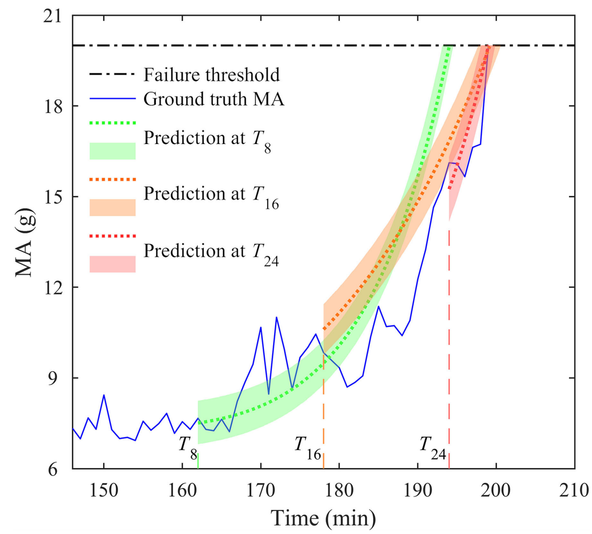


Fig. 7. Degradation predictions of Bearing 2_1 at inspection time $T_8 = 162\,\text{min}$, $T_{16} = 178\,\text{min}$, and $T_{24} = 194\,\text{min}$.

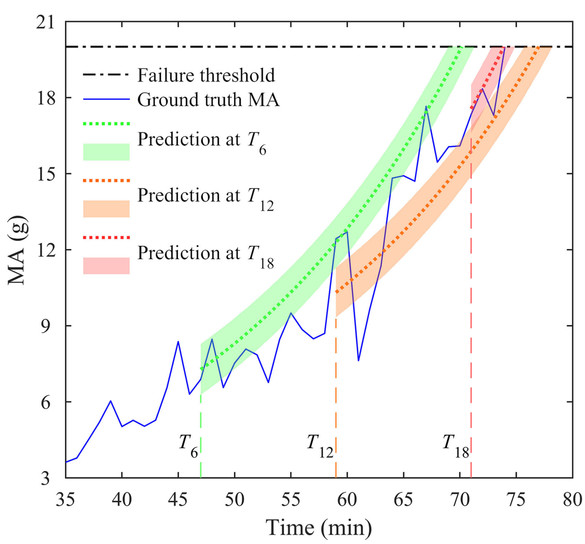


Fig. 6. Degradation predictions of Bearing 1_1 at inspection time $T_6 = 47\,\text{min}$, $T_{12} = 59\,\text{min}$, and $T_{18} = 71\,\text{min}$.

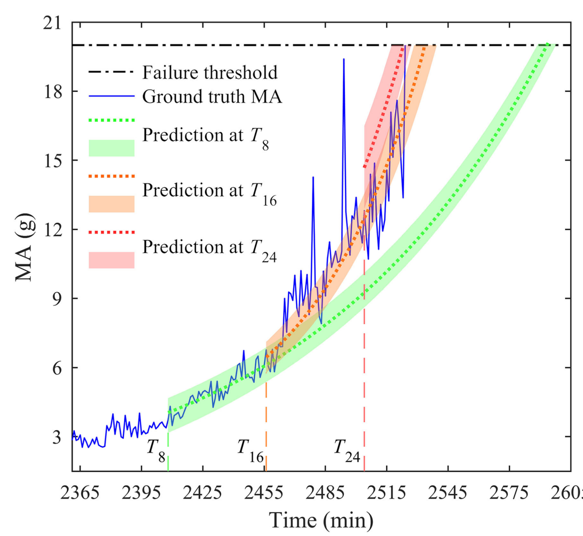


Fig. 8. Degradation predictions of Bearing 3_1 at inspection time $T_8 = 2409\,\text{min}$, $T_{16} = 2457\,\text{min}$, and $T_{24} = 2505\,\text{min}$.

curves, the optimal degradation curve can be obtained, which is plotted by orange solid line in Fig. 5. Subsequently, the MA values at the future time steps are predicted by extrapolating the selected optimal fitted curve. When the predictive MA value first hits the predefined failure threshold, i.e., 20 g, Bearing 3_1 is considered to be unusable, and the RUL at T_{14}, i.e., $\text{RUL}(T_{14})$, can be computed according to (16). Moreover, the uncertainty boundaries of the degradation prediction, which are depicted by the orange dashed lines, are also got via using (14) to fit the 95% upper and lower confidence limits of the selected RVs, respectively. Therefore, as shown in Fig. 5, the upper and lower boundaries of $\text{RUL}(T_{14})$ can be separately calculated by (16) when the corresponding uncertainty boundary is extrapolated to reach the failure threshold.

Further, Figs. 6–8 show the degradation prediction results of Bearing 1_1, Bearing 2_1, and Bearing 3_1 at three different inspection times, respectively. The predictive MA values obtained by extrapolating the optimal fitted curve are shown by the dotted lines, and the corresponding uncertainty interval is indicated by the shaded regions. It can be seen that the prediction at the initial degradation stage fails to follow the actual bearing degradation trend, and hence, the predictive failure time is earlier or later than the actual failure time. However, when more degradation data are available, the prediction is able to pick up the degradation trend fairly well, and so the predictive failure time gets closer to the actual one. Moreover, it is noticed that the uncertainty interval of the prediction is very narrow, which leads to the low uncertainty of RUL estimation. By calculating the time interval between inspection time and predictive failure time, the RUL of the tested bearings at each inspection time can be estimated. Figs. 9–11 present the RUL estimation results of Bearing 1_1, Bearing 2_1, and Bearing 3_1, respectively.

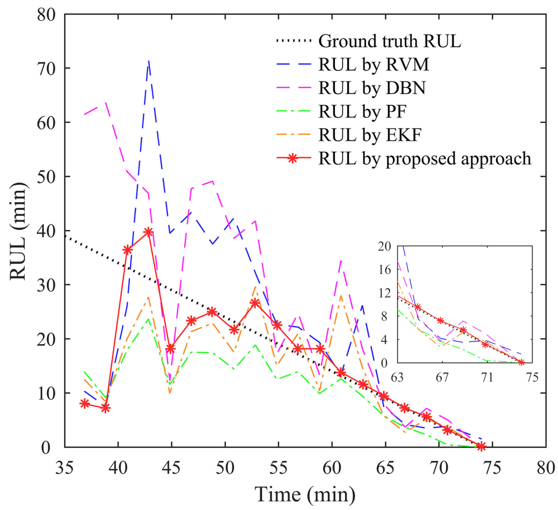


Fig. 9. RUL estimation results using five different prognostics approaches for Bearing 1_1.

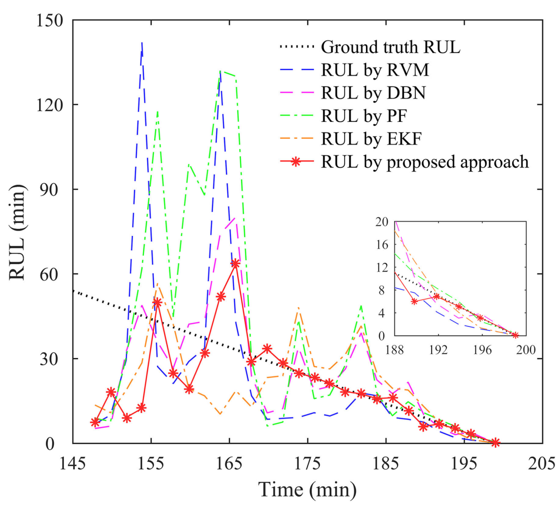


Fig. 10. RUL estimation results using five different prognostics approaches for Bearing 2_1.

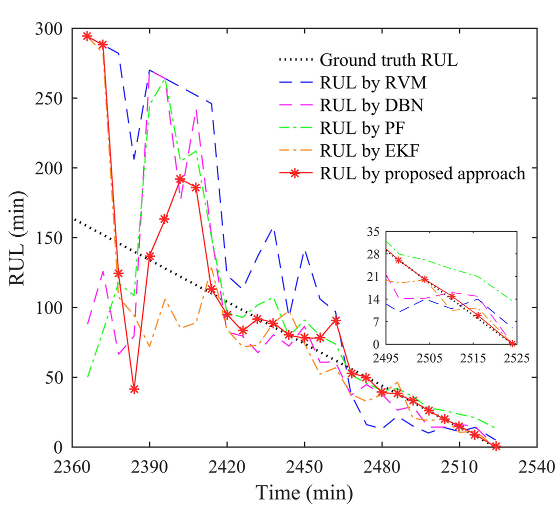


Fig. 11. RUL estimation results using five different prognostics approaches for Bearing 3_1.

To illustrate the superiority of the proposed hybrid prognostics approach compared with the purely data-driven approaches and the purely model-based approaches, two commonly used data-driven approaches, i.e., RVM-based approach [15] and deep belief network (DBN) based approach [11], and two commonly used model-based approaches, i.e., PF-based approach [23] and EKF-based approach [6], are also implemented to estimate the RUL of the tested bearings at each inspection time. For a fair comparison, all of these prognostics approaches use MA as the prognostics feature. Correspondingly, the RUL estimation results of Bearing 1_1, Bearing 2_1, and Bearing 3_1 are also shown in Figs. 9–11, respectively. It can be observed that the RUL estimation results of these five prognostics approaches all deviate from the actual RUL at the beginning, but as time goes on, the deviation reduces gradually and the estimated RUL converges to the actual RUL. The proposed hybrid prognostics approach, however, yields faster convergence and more accurate RUL estimation results in comparison to the other four prognostics approaches. The performance improvement of the hybrid approach can be attributed to the following two aspects.

1) The hybrid approach involves the prior knowledge of bearing degradation, which effectively reduces RUL estimation errors, especially when a few degradation data are available. In contrast, the purely data-driven approaches, i.e., both RVM-based and DBN-based approaches, have very large prediction errors in the absence of sufficient degradation data and their convergence is slower than that of the hybrid approach. Additionally, when the end-of-life approaches, the hybrid approach also presents more accurate prediction results than the two purely data-driven approaches.
2) In PF-based and EKF-based approaches, the unknown model parameters need to be initialized by a prior distribution before estimating them using PF or EKF. However, the appropriate selection of the prior distributions requires explicit prior knowledge or extensive empirical data, which is full of challenges in most prognostics applications. Moreover, it should be noted that PF often suffers from the particle degeneracy problem. The hybrid approach, by contrast, is able to directly estimate the unknown model parameters using the different RVs obtained by RVM regressions, and then employs the Fréchet distance to find the optimal one. Therefore, the hybrid approach achieves higher RUL estimation accuracy and better generalization.

To evaluate the prediction performance of these five prognostics approaches quantitatively, two widely used evaluation metrics, i.e., cumulative relative accuracy (CRA) and convergence [38], are employed in this paper. CRA is able to comprehensively assess the accuracy of a prognostics approach by aggregating the relative prediction accuracies at all inspection time. Given RUL prediction results, CRA value is calculated by

$$\mathrm{CRA} = \sum_{k=1}^{K} w_k \mathrm{RA}\left(T_k\right) \tag{17}$$

TABLE II
PERFORMANCE EVALUATION FOR ALL FIVE PROGNOSTICS APPROACHES USING TWO METRICS

Operating condition	Bearing dataset	RVM		DBN		PF		EKF		Proposed approach	
		CRA	C_{PE}	CRA	C_{PE}	CRA	C_{PE}	CRA	C_{PE}	CRA	C_{PE}
Condition 1	Bearing 1_1	0.5741	17.5669	0.4318	18.3295	0.6107	16.3494	0.6209	16.3386	0.9047	15.6174
	Bearing 1_2	0.1815	5.4003	0.6248	4.5782	0.7256	4.4768	0.3500	5.2880	0.8546	4.8239
	Bearing 1_3	0.6245	39.6812	0.5571	39.2412	0.4850	40.2539	0.8010	38.2853	0.8482	36.6895
	Bearing 1_4	0.3722	31.6046	-0.9479	31.3866	0.2305	29.5439	0.6839	27.1754	0.7240	26.7509
	Bearing 1_5	0.6122	5.3522	0.6636	4.7856	0.4311	5.1985	0.5042	5.0384	0.7878	4.9110
Condition 2	Bearing 2_1	0.5718	27.2026	0.5518	24.5515	0.3963	30.9560	0.5150	24.7948	0.8621	23.6716
	Bearing 2_2	0.1789	43.8837	-0.1977	46.5831	0.2634	48.6846	0.4314	46.9580	0.6521	40.1065
	Bearing 2_3	0.6172	11.6119	0.9013	10.6858	0.7364	11.2617	0.8800	10.6518	0.9612	10.5418
	Bearing 2_4	-0.0693	135.9779	0.5316	131.7427	0.4633	127.5825	0.5004	126.7525	0.6276	113.9724
	Bearing 2_5	0.2563	18.0174	0.0671	20.4009	0.1833	19.8443	0.4815	19.0677	0.6328	18.3933
Condition 3	Bearing 3_1	0.4329	86.8858	0.6979	75.3254	0.6557	73.9843	0.7744	72.0947	0.8942	72.0511
	Bearing 3_2	0.2225	3.4897	-0.1301	3.7609	0.1518	3.2065	0.5362	3.1503	0.6517	2.5632
	Bearing 3_3	-0.0883	95.7748	0.5167	61.2214	0.1283	69.4605	-0.9653	92.4687	0.8887	55.5899
	Bearing 3_4	0.6570	15.5349	0.6050	12.4770	0.7830	12.1350	0.6670	13.6413	0.8133	11.8487
	Bearing 3_5	0.4881	33.6375	0.5618	30.2354	0.3857	32.2596	0.5040	32.8719	0.6512	28.5316

where w_k is a normalized weight factor with $w_k = k / \sum_{k=1}^{K} k$, and $\mathrm{RA}(T_k)$ is the relative prediction accuracy at inspection time T_k

$$\mathrm{RA}\,(T_k) = 1 - \frac{|\mathrm{ActRUL}\,(T_k) - \mathrm{RUL}(T_k)|}{\mathrm{ActRUL}\,(T_k)} \qquad (18)$$

where $\mathrm{ActRUL}(T_k)$ is the actual RUL at inspection time T_k and $\mathrm{RUL}(T_k)$ is given by (16). The closer CRA value is to 1, the more accurate the RUL estimation results of the prognostics approach are.

Convergence is able to measure the speed that the estimated RUL converges to the actual RUL. It is defined as the Euclidean distance between the origin and the centroid of the area under the prediction error curve given by

$$C_{\mathrm{PE}} = \sqrt{(C_x - T_1)^2 + {C_y}^2} \qquad (19)$$

where T_1 is the first inspection time and (C_x, C_y) is the centroid of the area under the prediction error curve. The lower convergence value implies that the prediction results converge faster to the actual RUL as more degradation information is acquired over time.

Table II summarizes the performance evaluation results of these five prognostics approaches. From this table, it is clearly seen that the proposed approach has higher CRA value and lower convergence value than the other four approaches for each of the tested bearings, which means that the developed hybrid prognostics strategy is able to provide more accurate RUL estimation results and achieve faster convergence. Consequently, the proposed approach is superior to the other prognostics approaches in RUL prediction of rolling element bearings.

V. CASE STUDY II: PUBLIC PRONOSTIA BEARING DATASETS

To further verify the effectiveness and superiority of the proposed hybrid prognostics approach, a group of public bearing datasets, i.e., PRONOSTIA bearing datasets [39], is used in this section. PRONOSTIA bearing datasets are provided by the Franche-Comté Electronics Mechanics Thermal Science and Optics-Science and Technologies Institute and are utilized as the IEEE PHM 2012 Data Challenge problem. As tabulated in Table III, the datasets consist of three different operating conditions, and a total of seventeen run-to-failure datasets are given, including six training datasets and eleven testing datasets. In particular, the monitoring data of the eleven testing bearings are truncated so as to predict their RUL.

As prognostics benchmark datasets, PRONOSTIA bearing datasets have been widely used in the field of RUL prediction. Therefore, in order to validate the superiority of the proposed hybrid prognostics approach in RUL prediction of bearings, two published studies are employed in this section for comparison. These two studies were presented by Hong *et al.* [40] and Liu *et al.* [41], respectively, and both of them used PRONOSTIA bearing datasets as validation datasets. In [40], a fusion feature was constructed by combining wavelet packet decomposition, empirical mode decomposition and self-organizing map to reflect health states of bearings, and then a purely data-driven prognostics approach, i.e., Gaussian process regression, was used to estimate the RUL. Whereas

TABLE III
OPERATING CONDITIONS OF PRONOSTIA BEARING DATASETS

Operating condition	Radial force (N)	Rotating speed (rpm)	Training dataset	Testing dataset
Condition A	4000	1800	Bearing A_1 Bearing A_2	Bearing A_3 Bearing A_4 Bearing A_5 Bearing A_6 Bearing A_7
Condition B	4200	1650	Bearing B_1 Bearing B_2	Bearing B_3 Bearing B_4 Bearing B_5 Bearing B_6 Bearing B_7
Condition C	5000	1500	Bearing C_1 Bearing C_2	Bearing C_3

TABLE IV
PERFORMANCE COMPARISON OF THREE DIFFERENT PROGNOSTICS APPROACHES ON PRONOSTIA BEARING DATASETS

Testing dataset	Inspection time (s)	Actual RUL (s)	Estimated RUL 1 (s)	Estimated RUL 2 (s)	Er 1 (%)	Er of [40] (%)	Er 2 (%)	Er of [41] (%)
Bearing A_3	18010	5730	5790	5440	-1.05	-1.04	5.06	2.58
Bearing A_4	11380	339	400	260	-17.99	-20.94	23.30	-9.14
Bearing A_5	23010	1610	1260	1540	21.74	-278.26	4.35	-0.99
Bearing A_6	23010	1460	1370	1450	6.16	19.18	0.68	6.03
Bearing A_7	15010	7570	6980	10790	7.79	-7.13	-42.54	-0.70
Bearing B_3	12010	7530	4290	6220	43.03	10.49	17.40	55.44
Bearing B_4	6110	1390	1370	1220	1.44	51.80	12.23	15.56
Bearing B_5	20010	3090	2510	3100	18.77	28.80	-0.32	49.19
Bearing B_6	5710	1290	1260	1320	2.33	-20.93	-2.33	38.53
Bearing B_7	1710	580	600	530	-3.45	44.83	8.62	5.17
Bearing C_3	3510	820	710	850	13.41	-3.66	-3.66	2.56
Score					0.6234	0.3551	0.6682	0.6101
SD of Er					15.8955	90.2924	17.0173	22.2020

in [41], Hilbert–Huang transform was used to extract the prognostics features from raw vibration signals, and then a purely model-based prognostics approach, i.e., PF-based approach, was utilized to predict RUL of bearings. Accordingly, for a fair comparison, the features used in [40] and [41] are first extracted from vibration signals of bearings. After that, the proposed hybrid prognostics approach is employed to predict RUL of bearings at predefined inspection time.

Table IV reports the RUL prediction results of eleven testing datasets. The estimated RUL values using the features from [40] and [41] are denoted as RUL 1 and RUL 2, respectively. To comprehensively compare the performance of the proposed approach and the two approaches published in [40] and [41], a scoring function used in the IEEE PHM 2012 Data Challenge is employed in this section, which is defined as follows:

$$\text{Score} = \frac{1}{11}\sum_{i=1}^{11} A_i \tag{20}$$

where

$$A_i = \begin{cases} \exp\left(-\ln(0.5)\cdot(Er_i/5)\right) & \text{if } Er_i \le 0 \\ \exp\left(+\ln(0.5)\cdot(Er_i/20)\right) & \text{if } Er_i > 0 \end{cases} \tag{21}$$

and Er_i is the percent error of RUL prediction results for ith testing dataset and can be calculated by the following equation:

$$Er_i = \frac{\text{ActRUL}_i - \text{RUL}_i}{\text{ActRUL}_i} \times 100\%. \tag{22}$$

The percent errors of RUL 1 and RUL 2 are given in the columns of Er 1 and Er 2 in Table IV, respectively. Correspondingly, the percent errors of [40] and [41] are also listed in Table IV. According to (20) and (21), the scores of these three prognostics approaches are, respectively, calculated and shown at the bottom of Table IV. Meanwhile, in order to evaluate the stability of RUL prediction results, the standard deviation (SD) of the percent errors is computed as well. From Table IV, it can be observed that the proposed hybrid prognostics approach achieves higher score compared with the other two approaches, and its SD is also smaller. This implies that the proposed hybrid prognostics approach provides more accurate and more stable RUL estimation results in RUL prediction of rolling element bearings.

VI. CONCLUSION

In this paper, a hybrid prognostics approach was proposed for estimating RUL of rolling element bearings. The proposed approach consists of three major processes: sparse representations of degradation data; degradation curve identification; and RUL prediction. First, the available degradation data, i.e., the features extracted from degradation signals, were sparsely represented using RVM regressions with different kernel parameters. Then, the single exponential function and the sum of two exponential functions were used to fit each RVs obtained by RVM regressions, and the Fréchet distance is employed to select the optimal degradation curve from all fitted curves. Finally, the RUL was predicted by extrapolating the optimal degradation curve to reach the failure threshold.

The proposed hybrid prognostics approach integrates the advantages of both data-driven and model-based approaches while evading their respective limitations. On one hand, RVM regression reduces the interference of measurement noise and removes irrelevant and redundant degradation data. More importantly, through setting different kernel parameters, the sparsity of RVM can be made of full use, and thus, various kinds of representative degradation information can be obtained from the limited measurements. On the other hand, the hybrid approach directly estimates the unknown parameters of the exponential degradation models by fitting different RVs and then selects the optimal one with the help of the Fréchet distance, thus, effectively alleviating the problem of model parameters initialization. Simultaneously, the combination of the Fréchet distance and the exponential degradation models is able to better describe the possible degradation paths of a bearing, and so the degradation predictions are more accurate.

The proposed approach was first evaluated using the vibration data acquired from the accelerated degradation tests of rolling element bearings, and also compared with four state-of-the-art prognostics approaches, including two purely data-driven approaches, i.e., RVM-based approach and DBN-based approach, and two purely model-based approaches, i.e., PF-based approach and EKF-based approach. Experimental results demonstrated that the developed hybrid prognostics approach is superior to the other four existing prognostics approaches in RUL prediction of bearings. Further, the public PRONOSTIA bearing datasets were also used to validate the effectiveness and superiority of the proposed approach, and the results showed that the proposed hybrid prognostics approach achieves more accurate and more stable RUL estimation results in comparison with the two published studies.

It should be mentioned that the selection of the degradation models is very flexible in the proposed hybrid prognostics approach. That is, based on the specific applications, the exponential functions used in this paper can be replaced by some other basic degradation functions, such as the polynomial functions. Therefore, the proposed approach can be easily extended to deal with the degradation problems of some key machine components, such as gears, shafts, and cutting tools. Moreover, although the sequential sparse Bayesian learning algorithm is used to optimize the hyperparameters of the RVM model in this paper, the computational overhead caused by RVM regressions still gradually increases as more degradation data are acquired, which may pose a limitation for the proposed approach in practical applications. Consequently, in the future research, a more effective hyperparameter optimization algorithm will be developed to improve the training speed of RVM regression on large datasets.

References

[1] M. Cerrada *et al.*, "A review on data-driven fault severity assessment in rolling bearings," *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, 2018.

[2] N. H. Kim, D. An, and J.-H. Choi, *Prognostics and Health Management of Engineering Systems: An Introduction*: Berlin, Germany: Springer, 2017.

[3] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.

[4] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *Adv. Mech. Eng.*, vol. 8, pp. 1–21, 2016.

[5] H. Liao and Z. Tian, "A framework for predicting the remaining useful life of a single unit under time varying operating conditions," *IIE Trans.*, vol. 45, pp. 964–980, 2013.

[6] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.

[7] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018.

[8] K. L. Tsui, N. C. Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Math. Probl. Eng.*, vol. 2015, pp. 1–17, 2015.

[9] X. Si, W. Wang, C. Hu, and D. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, pp. 1–14, 2011.

[10] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.

[11] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.

[12] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.

[13] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.

[14] F. Di Maio, K. L. Tsui, and E. Zio, "Combining relevance vector machines and exponential regression for bearing residual life estimation," *Mech. Syst. Signal Process.*, vol. 31, pp. 405–427, 2012.

[15] Y. Zhang and B. Guo, "Online capacity estimation of lithium-ion batteries based on novel feature extraction and adaptive multi-kernel relevance vector machine," *Energies*, vol. 8, pp. 12439–12457, 2015.

[16] A. Soualhi, H. Razik, and G. Clerc, "Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2864–2874, Jul. 2014.

[17] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vol. 56/57, pp. 150–172, 2015.

[18] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 647–656, Jan. 2015.

[19] L. Liao and F. Köttig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction," *IEEE Trans. Rel.*, vol. 63, no. 1, pp. 191–207, Mar. 2014.

[20] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K. L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Inform.*, vol. 12, no. 3, pp. 924–932, Jun. 2016.

[21] Y. Qian, R. Yan, and S. Hu, "Bearing degradation evaluation using recurrence quantification analysis and Kalman filter," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2599–2610, Nov. 2014.

[22] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. S. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 2046–2054, Sep. 2016.

[23] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2464–2472, May 2014.

[24] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.

[25] N. Bolander, H. Qiu, N. Eklund, E. Hindle, and T. Rosenfeld, "Physics-based remaining useful life prediction for aircraft engine bearing prognosis," *presented at the Annu. Conf. Prognostics Health Manage. Soc.*, San Diego, CA, USA, 2009.

[26] R. Moghaddass and M. Zuo, “A parameter estimation method for a condition-monitored device under multi-state deterioration,” *Rel. Eng. Syst. Saf.*, vol. 106, pp. 94–103, 2012.

[27] J. Son, Q. Zhou, S. Zhou, and M. Salman, “Prediction of the failure interval with maximum power based on the remaining useful life distribution,” *IIE Trans.*, vol. 47, pp. 1072–1087, 2015.

[28] J. Son, Y. Zhang, C. Sankavaram, and S. Zhou, “RUL prediction for individual units based on condition monitoring signals with a change point,” *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 182–196, Mar. 2015.

[29] Y. Chen, L. Yang, C. Ye, and R. Kang, “Failure mechanism dependence and reliability evaluation of non-repairable system,” *Rel. Eng. Syst. Saf.*, vol. 138, pp. 273–283, 2015.

[30] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.

[31] M. E. Tipping, “Fast marginal likelihood maximisation for sparse Bayesian models,” *presented at the Proc. 9th Int. Workshop Artif. Intell. Statist.*, Key West, FL, USA, 2003, pp. 1–13.

[32] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method,” *J. Power Sources*, vol. 196, pp. 10314–10321, 2011.

[33] T. Strutz, *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*, 1st ed. Wiesbaden, Germany: Vieweg and Teubner, 2011.

[34] J. Iwaniec, W. Lisowski, and T. Uhl, “Nonparametric approach to improvement of quality of modal parameters estimation,” *J. Theor. Appl. Mech.*, vol. 43, pp. 327–344, 2005.

[35] S. Har-Peled and B. Raichel, “The Fréchet distance revisited and extended,” *ACM Trans. Algorithms*, vol. 10, Jan. 2014, Art. no. 3.

[36] A. Efrat, L. J. Guibas, S. Har-Peled, and J. S. B. Mitchell, “New similarity measures between polylines with applications to morphing and polygon sweeping,” *Discrete Comput. Geom.*, vol. 28, pp. 535–569, 2002.

[37] X. Si, W. Wang, C. Hu, M.-Y. Chen, and D.-H. Zhou, “A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation,” *Mech. Syst. Signal Process.*, vol. 35, pp. 219–237, 2013.

[38] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “Metrics for offline evaluation of prognostic performance,” *Int. J. Prognostics Health Manage.*, vol. 1, pp. 4–23, Jan. 2010.

[39] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, “PRONOSTIA: An experimental platform for bearings accelerated degradation tests,” *presented at the IEEE Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, 2012, pp. 1–8.

[40] S. Hong, Z. Zhou, E. Zio, and K. Hong, “Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method,” *Digit. Signal Process.*, vol. 27, pp. 159–166, 2014.

[41] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, “Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1549–1558, Feb. 2018.

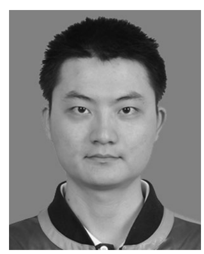
Biao Wang (S’17) received the B.S. and M.S. degrees in aeronautical and astronautical science and technology in 2014 and 2016, respectively from Xi’an Jiaotong University, Xi’an, China, where he is currently working toward the Ph.D. degree in mechanical engineering with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System.

His research interests include machinery condition monitoring, intelligent fault diagnostics, and remaining useful life prediction of rotating machinery.

Yaguo Lei (M’15) received the B.S. and Ph.D. degrees in mechanical engineering from Xi’an Jiaotong University, Xi’an, China, in 2002 and 2007, respectively.

He is currently a Full Professor of Mechanical Engineering with Xi’an Jiaotong University. Prior to joining Xi’an Jiaotong University in 2010, he was a Postdoctoral Research Fellow with the University of Alberta, Edmonton, AB, Canada. He was also an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Duisburg, Germany. He has pioneered many signal processing techniques, intelligent diagnosis methods, and remaining useful life prediction models for machinery. His research interests include machinery condition monitoring and fault diagnosis, mechanical signal processing, intelligent fault diagnostics, and remaining useful life prediction.

Dr. Lei is a member of the Editorial Boards of more than ten journals, including *Mechanical System and Signal Processing* and *Neural Computing & Applications*. He is also a member of the ASME.



Naipeng Li received the B.S. degree in mechanical engineering from Shandong Agricultural University, Tai’an, China, in 2012. He is currently working toward the Ph.D. degree in mechanical engineering with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi’an Jiaotong University, Xi’an, China.

His research interests include machinery condition monitoring, intelligent fault diagnostics, and remaining useful life prediction of rotating machinery.

Ningbo Li received the B.S. degree in mechanical engineering from Sichuan University, Chengdu, China, in 2015. He is currently working toward the M.S. degree in mechanical engineering with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi’an Jiaotong University, Xi’an, China.

His research interests include machinery condition monitoring, intelligent fault diagnostics, and remaining useful life prediction of rotating machinery.